DOCUMENT RESUME

ED 429 107 TM 029 646

AUTHOR	Butler, Frances A.; Stevens, Robin
TITLE	Initial Steps in the Validation of the Second Language Proficiency Descriptors for Public High Schools, Colleges, and Universities in California: Writing.
INSTITUTION	Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.; California Community Colleges, Sacramento. Office of the Chancellor.
REPORT NO	CSE-TR-497
PUB DATE	1998-10-00
NOTE	53p.
CONTRACT	R305B60002
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Administrators; College Students; High School Students; High Schools; Higher Education; Public Colleges; Public Schools; *Second Language Learning; State Universities; *Subject Index Terms; Teachers; *Validity; *Writing (Composition)
IDENTIFIERS	California; Experts

ABSTRACT

The need for empirical validation of a specific set of second language proficiency descriptors for the four skill areas- -reading, writing, listening, and speaking--provided the impetus for the work described in this report. The University of California at Los Angeles Center for the Study of Evaluation developed a validation plan and undertook initial steps in the validation process with one skill area, writing. The process, which includes anchoring descriptor levels to student performance, involved the participation of writing experts from high schools, colleges, and universities across California. In addition, nine potential descriptor users from the same educational segments across the state were asked to help clarify descriptor applications. Work with the writing descriptors led to refinement of the validation process. The report includes a detailed description of that process and provides suggestions for steps that can be taken to validate the descriptors for reading, listening, and speaking. Eight appendixes contain lists of participants, the user interview protocol, descriptions of proficiency descriptors, writing samples, and some worksheets for the study. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

ED 429 107

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Kim Hurst
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

Initial Steps in the Validation of the Second Language Proficiency Descriptors for Public High Schools, Colleges, and Universities in California: Writing

CSE Technical Report 497

Frances A. Butler

Robin Stevens

Center for the Study of Evaluation
University of California, Los Angeles

TM029646

BEST COPY AVAILABLE

Initial Steps in the Validation of the Second Language Proficiency Descriptors for Public High Schools, Colleges, and Universities in California: Writing

CSE Technical Report 497

Frances A. Butler

Robin Stevens

Center for the Study of Evaluation
University of California, Los Angeles

October 1998

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Copyright © 1999 The Regents of the University of California

The work reported herein was supported in part by Orange Coast College through a joint faculty project grant from the Chancellor's Office of the California Community Colleges and by the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U. S. Department of Education.

The opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the positions or policies of Orange Coast College or the Chancellor's Office of the California Community Colleges, nor do they reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U. S. Department of Education.

INITIAL STEPS IN THE VALIDATION OF THE SECOND LANGUAGE PROFICIENCY DESCRIPTORS FOR PUBLIC HIGH SCHOOLS, COLLEGES, AND UNIVERSITIES IN CALIFORNIA: WRITING[1]

Frances A. Butler

Robin Stevens

Center for the Study of Evaluation
University of California, Los Angeles

ABSTRACT

The need for empirical validation of a specific set of second language proficiency descriptors for the four skill areas—reading, writing, listening, and speaking—provided the impetus for the work described in this report. The UCLA Center for the Study of Evaluation developed a validation plan and undertook initial steps in the validation process with one skill area, writing. The process, which includes anchoring descriptor levels to student performance, involved the participation of writing experts from high schools, colleges, and universities across California. In addition, potential descriptor users from the same educational segments across the state were asked to help clarify descriptor applications. Work with the writing descriptors led to refinement of the validation process. The report includes a detailed description of that process and provides suggestions for steps that can be taken to validate the descriptors for reading, listening, and speaking.

[1] The support of many made the undertaking described in this report possible. Students, teachers, and site coordinators at high school, community college, California State University, and University of California campuses across the state willingly participated in this effort. The working group members of the English as a Second Language (ESL) Intersegmental Project made initial contacts at the campuses and provided additional contacts and other types of assistance as needed. Gari Browning and Julie Thornton read earlier drafts of this report and provided valuable comments. Judy Miyoshi at the UCLA Center for the Study of Evaluation provided administrative support and assisted with the organization of data collection. Katie Hutton, a student worker at UCLA, provided cheerful assistance to project staff. To all we express our sincere gratitude and appreciation.

We wish to express a special *thank you* to Jean Turner, a specialist in language testing at the Monterey Institute of International Studies, who served as advisor to the project. Her input and guidance at every stage contributed in large measure to our progress throughout. She helped shape the data analysis and provided insightful feedback on earlier drafts of this report.

Finally, we dedicate this work to the ESL students in California who face numerous linguistic challenges in their quest for an education. It is our hope that the effort described here will in some way facilitate their educational journeys.

Introduction

The study described in this report is the first step in a validation effort for the second language proficiency descriptors in *California Pathways: The Second Language Student in Public High Schools, Colleges, and Universities*.[2] The goal of this work is to empirically validate theoretically and experientially derived descriptors by characterizing the performance levels of writing proficiency through analysis of student writing samples. The current focus is on the writing descriptors with the intention that progress in this skill area will inform the validation process for use with the other skills.

The descriptors were developed for four language skills—reading, writing, listening, and speaking. They are intended to capture the full continuum of second language proficiency from rank beginner to learners who are indistinguishable from native speakers and are for use within and across all segments of the education system in California, from high school through college. More specifically, the descriptors give those who work with second language (L2) learners a common language to approach the following areas:

- discussing the continuum of L2 proficiency levels;
- developing or revising ESL curricula;
- evaluating [and developing] tests;
- interpreting courses within and across segments. (*California Pathways*, p. 77)

This range of potential uses highlights important application possibilities for the descriptors and underscores the need for assuring accuracy and validity of the descriptors vis à vis actual student performance. A plan for validating the descriptors was developed which provides an approach to anchoring the descriptor levels to student performance (Butler & Stevens, 1997).

Since guidelines do not exist for validating language proficiency descriptors of this kind, a major part of the work described in the plan involved developing and refining a validation process that can be used for the descriptors from all

[2] *California Pathways: The Second Language Student in Public High Schools, Colleges, and Universities*, henceforth in this report referred to as *California Pathways*, was written in 1995 by ESL Intersegmental Project members with funding from the California Community College Chancellor's Office, Intersegmental Joint Faculty Project.

four skill areas, noting that modifications may be needed for some steps in the process due to differences in modalities. For example, listening and reading performance will be more difficult to capture because these skills cannot be observed in isolation from others. It may be necessary to use established tests to help tap listening and reading ability. Irrespective of how performance is captured, to validate the descriptors, samples of language performance must be obtained for each skill area.

In order to develop and ultimately conduct the validation process for the descriptors from all of the skill areas as efficiently as possible, initial efforts have focused on writing. Writing was selected because it is often such a high-stakes skill for English as a second language (ESL) students, in terms of entrance to educational institutions, in terms of placement decisions, and for their own success. This report describes the steps taken in the validation process for the writing descriptors and, based on the findings of the study, provides suggestions for next steps toward final validation of the writing descriptors, as well as steps that can be taken to validate the descriptors for reading, listening, and speaking.

The validation study required that staff from the UCLA Center for the Study of Evaluation (CSE) be granted access to students and professionals across the four segments cited above. To assure this access, the same intersegmental project members who developed *California Pathways* and the language proficiency descriptors agreed to serve as a working group to provide support for specific validation tasks (see Appendix A for the members of the 1997-1998 working group).

Validation Issues

California Pathways was the guiding document for the validation plan and thus also for the work described in this report. It provided the initial foundation for the validation study through general discussion of how the proficiency descriptors might be used.

> [The second language proficiency descriptors] give ESL specialists and others who are in contact with this population a way to connect the language education paths of a significant portion of California's students. (*California Pathways*, p. xxi)

Because a number of stakeholder groups may use the descriptors, it was critical to develop a better understanding of who those users are and what the descriptor

applications will be. *California Pathways* implies there may be multiple types of users from both inside and outside of the ESL field with a potentially wide range of needs; thus it was also critical to determine to what degree those users can apply the descriptors as currently configured and to what degree, if any, the language and structure of the descriptors need to be revisited.

Another important consideration in the user-descriptor picture is the comparability of the performance levels as currently described for all segments—high school, community college, California State University (CSU), and the University of California (UC). That is, do the descriptor levels carry the same meaning across segments and can the distinctions drawn by the current number of levels be identified clearly in student performance for the purposes intended by the users?

These validation issues helped to shape the methods used for conducting this study. The work carried out to date on validating the writing descriptors provides some answers to these questions, including how the descriptors might be used, who has a need for them, and how these two considerations should shape the language and organization of the descriptors.

The Validation Process for Writing

The process for validating the writing descriptors included interviews as well as traditional empirical methods. End users were interviewed to help clarify how the writing descriptors might be used across segments. Next four promising writing tasks were piloted with English language learners and native speakers at one school per segment which led to the selection of the two best tasks for use in a larger sample collection effort. Writing samples were then collected from students in all four segments across the proficiency range of beginning to advanced. A representative sample of papers was selected from the full range of classes and schools that participated. Two groups of ESL writing experts were asked to sort the selected writing samples into proficiency levels based on what they perceived to be the distinguishing differences in the quality of writing across the samples and to identify and describe the critical language features that define the levels established by the group. Exemplar papers were also identified for some of the levels by one of the groups. CSE staff compiled the results and then compared the levels and language features to the writing descriptors to help determine the appropriateness of the existing levels and the accuracy of the

language in the descriptors vis à vis actual samples of writing from the population to be served. A working group subcommittee sorted a set of the same writing samples into descriptor levels to provide additional information regarding the applicability of the descriptors to actual samples of student work. Finally, end users were asked to read a set of writing samples and assign each paper to one of the writing descriptor levels. The initial plan for each step in this process is discussed in Butler and Stevens (1997). The outcomes and findings for each step are discussed in turn below.

End-user Interviews

Clarification of descriptor applications has been an on-going process in this validation study and will continue to be in future efforts since, in order for the descriptors to be valid for particular uses, those uses need to be clearly articulated. To this end, nine potential end users from across the four segments were interviewed. Working group members helped to identify the participants. There were two participants each from the high school, community college, and CSU segments, and three from the UC segment. Four of the participants are currently teaching ESL students in addition to their other job duties. The others all fill administrative and counseling roles at their institutions (see Appendix B for a list of the participant job titles and schools). Each interview lasted approximately thirty minutes. The interview protocol consisted of nineteen questions, including five general questions and fourteen optional questions that were asked only if they were relevant to the participants' jobs (see Appendix C for the end-user interview protocol). The questions were designed to elicit data about the interviewees' job responsibilities related to ESL students, how they might use the descriptors in their work, and what their needs are regarding descriptors in other skill areas. Three key areas that emerged from the interviews are discussed below: need for the descriptors, descriptor uses, and proficiency levels.

Need for the descriptors. Data from the interviews indicated a serious need for a common framework such as the descriptors that allows people in a variety of educational occupations to speak with one another and with ESL students about ESL issues and needs. However, the usefulness of the descriptors is limited without language samples that illustrate each descriptor level, especially for people who are not considered ESL experts. A common language illustrated by concrete examples is clearly needed so that everyone has the same point of reference for discussing ESL proficiency issues. As it is now, each potential user

comes to the descriptors with a slightly different perspective regarding what they mean and how they should be interpreted.

Descriptor uses. The data from the interviews also provided evidence of the need for a tool, such as the descriptors, for articulation purposes. Since articulation had been named as a critical need by working group members, having it raised during the interviews by end users highlighted its importance. Interview participants felt that the descriptors should be used to correlate courses and/or course levels at each institution. They thought that a grid or comparability matrix showing which courses or course levels at one segment correspond to courses or course levels at another would be very useful to them in their everyday work.

Interviewees also noted a need for a tool which would help them discuss with ESL students individual student proficiency in terms of the English language skills needed to perform adequately in mainstream classes and in work environments. Participants often work with students directly in an advisory capacity and sometimes talk with parents or teachers about student performance levels and areas for student improvement. This need to discuss ESL proficiency relative to the proficiency of native speakers may be a major difference between the descriptor uses for ESL "experts" and "non-experts." ESL experts may have a greater need to discuss ESL proficiency levels with respect to other ESL students while non-experts may need to discuss ESL proficiency levels with respect to the level of English needed to participate adequately in mainstream classes with native speakers of English.

Proficiency levels. A related issue raised during the interviews was the number of proficiency levels participants felt they would need for their work with ESL students. Four of the nine interviewees stated that between three and six levels are necessary. Two others did not specify the number of levels but felt finer distinctions are needed at lower levels of ability and fewer at the upper levels. One person felt that ten levels is adequate and the other two did not know how many levels would be needed. The uncertainty about the number of levels needed may, in part, reflect the different foci of the participants in working with ESL students. Because of differences in focus, a part of validating the descriptors for use by different populations may include establishing the number of levels that are necessary for specific "expert" and "non-expert" uses. The non-expert users may require fewer but wider bands of proficiency and/or special guidelines

for using the descriptors. It may be that these "bands" are the overarching range categories within which ESL experts will find the more specific subcategories needed for ESL purposes.

Identification of Writing Tasks and Collection of Writing Samples

To help validate the writing descriptors and anchor them to student performance, writing samples were collected across all four segments and levels of proficiency. To obtain the necessary samples of academic writing, existing writing tests, prompts, and tasks from each segment were reviewed for their potential effectiveness in allowing writers from a wide range of ability and from different segments to respond. Tasks that included reading material in addition to the directives, such as a paragraph or passage, were avoided to prevent comprehension problems with students who have differing levels of reading ability. A preliminary set of four tasks was selected and tried out with students across segments and levels to determine which two of the four would be most likely to produce the best range of writing samples during data collection for the validity study.

Writing samples were collected at one site per segment in four classes at each site—three ESL classes, beginning, intermediate, and advanced, and one English class that included native speakers and proficient non-native speakers. Approximately 300 samples were collected. Based on the results of the student writing obtained from the tasks, two of the four were selected for the larger sample-collection effort.

Minor modifications were made to the writing tasks and task directions based on the task trials. The two selected topics were the following:

Topic A

> Choose two important people in your life, such as a teacher and a friend, two friends, or a relative and a classmate. Write an essay in which you discuss how they are similar and how they are different. Give specific examples.

Topic B

> Write an essay in which you discuss some difficulties that teenagers have growing up. Explain your opinion and give specific examples.

Topic A is a comparison/contrast task; Topic B is an analytic expository task. Both topics are considered academic in that they require students to use functions such

as analyze, explain, and compare to fulfill the task. However, both topics also allow for students to draw upon personal experience to justify, support, or generalize through the use of narrative. As Mlynarczyk (1991) notes, many of the functions used in personal writing such as narrative are also necessary in academic writing. In fact, the distinction between the two may be fine. Given this, and the need to limit the reading load, these two tasks were judged appropriate.

Using these two tasks, approximately 660 samples (330 per task) were collected for the full proficiency range across segments, including native speakers. Data collection sites included one school from each segment in both Southern California and Northern California (see Table 1). Packets of materials

Table 1

Sample Collection Plan

	Site			
Class	HS	CC	CSU	UC
Southern CA sites	(Site 1)	(Site 2)	(Site 3)	(Site 4)
Beginning ESL	21	14	16	10
Intermediate ESL	25	19	18	22
Advanced ESL	14	24	23	20
English*	98	13	24	25
Northern CA sites	(Site 5)	(Site 6)	(Site 7)	(Site 8)
Beginning ESL	22	21	13	7
Intermediate ESL	17	24	11	11
Advanced ESL	16	23	16	12
English	30	21	22	16

Note. HS = High school; CC = Community college; CSU = California State University; UC = University of California. In addition to the beginning through advanced ESL students, approximately 20 students from English classes that included native speakers and proficient non-native speakers were tested at each site. Total N = 668.

were mailed to a teacher or an administrator at each school who helped facilitate the sample collection. Each packet contained instructions and a task administration protocol for the teacher, student task booklets, and pens. Within

each class, students were randomly assigned one of the two selected tasks. They were given their normal class period to write on the task, usually between 45 and 50 minutes. Participation in the task was voluntary to meet University of California human subjects criteria. If students chose not to participate in the study, their teachers gave them an alternate non-graded task to complete. In general, across all segments, students were willing to participate when the purpose of the task was explained. All of the task materials were returned to CSE where the writing samples were reviewed and a range of representative samples was selected for the ESL writing expert sort described below.

ESL Writing Expert Sort

To examine the distinguishing characteristics of the descriptor levels, an iterative sorting process was conducted with ESL writing experts, two representatives from each segment. A similar process is discussed in Upshur and Turner (1995). The sort was conducted over a two-day period during which the eight representatives were split into two groups of four with one segment representative in each group. The expert sort was important because it was through this process that salient features of student writing critical to distinguishing group differences were identified.

The first step in the process involved having the experts in each group impressionistically and independently sort the samples into three broad levels of proficiency: beginning, intermediate, and advanced. They were then asked, as a group, to identify and describe the critical language features that enabled them to sort the papers into the three broad levels. In other words, they had to explain their rationale for placing a given paper at a particular level. The group members then tried to reach consensus on the papers they had placed into the three categories. One group was able to do this and the other group reached consensus on some of the papers and then moved on to the next step.

The groups were then asked to sort the papers into narrower levels of proficiency and, again, to articulate the critical language features that guided their sorting. The two groups produced a list of features for each of the levels they identified through the writing samples. CSE staff compared the two lists to one another and then to the descriptors to determine the degree of match.

Comparison of features from expert sort. A comparison of the sets of features generated by the two groups of ESL writing experts, each working with a

set of papers on a different topic, shows a high degree of similarity with regard to those elements of writing the experts perceived to be most salient in allowing them to distinguish the samples they examined. While the language used by the experts to identify the features was by no means identical for the two groups, CSE staff reorganized the level characteristics under general categories, such as *communicative success* and *organization*, which allowed for systematizing the features across the groups (see Appendix D for the two groups' language features by levels that were used for the comparisons). In some instances, the language used by one group is more precise and clearly stated than the other for a particular characteristic; regardless, the parallels between the two sets are striking.

Expert Group 1 identified eight or nine writing proficiency levels through the sorting process—three each for the low and mid ranges and two or three for the high range. Expert Group 2 identified seven or eight levels—three for low, two or three for the mid range, and two for high. The mid range was the most problematic for both groups in terms of their being able to articulate clear differences among the writing samples.

The two sets of features generated by the writing experts include for each of the three major ranges (low, mid, high): (a) a bulleted list of features at the top of the page that can be identified in all papers in that range; and (b) an additional set of subcategories. Group 1 first generated its list of features for each major range, and then generated new lists of features for each subcategory, independent of the range bullets. Thus, there is some overlap between the Group 1 range bullets and subcategory features. However, for Group 2, the bulleted points serve as a summary overview of the major range, and the statements for the subcategories provide the specificity for differentiating among them. A comparison of the groups' bulleted language features for the three major ranges and the subcategories shows commonalties in terms of categories and emphasis on categories within a range. Individual discussions of the three ranges follow.

In the low range, there are two key similarities: (a) the importance of *organization*, *communicative success*, and *length* in differentiating between papers; and (b) the lack of effectiveness when the groups attempted to quantify features as opposed to describing what is seen in a writing sample. *Organization* seems to be the feature that provides the most differentiation among the subcategories for the low range. The notions of "pre-paragraph" (from Group 1) and "emergent paragraph structure and emergent essay structure" (from Group

2) capture the developing organizational writing skills from Low Low to Low Mid to Low High. These notions are also present in Group 1's *length* category. Categories such as *structure and vocabulary*, while present in both groups' lists of features, seem to be less important at this range. The notion *communicative success* appears as a main category and feature for both groups but is not effectively articulated; for example, Group 2's description of *communicative success* for Low Mid and Low High respectively are "uneven comprehensibility" and "minimally comprehensible." The attempts at capturing differences for this characteristic are based on quantifiable rather than qualitative differences and may be less clear for this reason.

For the mid range, two important similarities emerged: (a) the groups articulated features that fell into the same exact categories; and (b) *organization* was an important distinguishing feature for both groups. The groups identified features that fell into the same categories: *communicative success, length, organization, structure,* and *vocabulary*. In the low and high ranges, the groups identified features that fell into at least one or more categories that do not overlap. *Structure,* while present in both, is not clearly described in either set of features. The notion of *communicative success* appears to be important for both groups at the mid range with primary focus on the impact of errors on comprehension. However, *organization* is the feature that provides the most differentiation among subcategories for the mid range, though the distinctions are not always clear. For both groups, the use of examples as a feature of *organization* was a key factor in capturing differences in the writing samples, as was the writer's ability to maintain focus.

The language features and categories in the high range again share similarities: (a) the emphasis on *organization* as a way to differentiate papers; and (b) the greater elaboration of *communicative success* and *vocabulary* at this range. *Communicative success, organization, structure,* and *vocabulary* all appear as strong characteristics for both groups at the high range. The statements in general are more positive and descriptive of what is seen in a sample. Though one group divided the high range into two subcategories and the other divided it into two or three, neither group felt that more than three levels were necessary to characterize the features of the writing samples used in the sort.

The participants in both groups felt the results of this sorting procedure were preliminary. Each group expressed the need for additional time and

samples to assure confidence in the number of levels and to refine the wording of the language features. While the discussion below involves revisiting both the features generated by the ESL writing experts and the current writing proficiency descriptors, having identified areas of overlap and similarity as well as gaps and differences between the groups' features and levels in this section will facilitate that process. A comparison of the language features to the current descriptors for writing follow.

Comparison of the descriptors to the expert levels. A comparison of the descriptors to the features generated by Groups 1 and 2 offers a contrast in perspective with regard to the nature of the descriptive information provided in each (see Appendix E for the reorganized version of the descriptors used for the comparison). While there is a high degree of overlap in the actual categories used in the descriptors and the two sets of writing features, especially at the intermediate (mid) and advanced (high) levels, there is an important difference in how the features are characterized. In general, the descriptors tend to quantify language features using qualifiers such as *some, often,* and *rarely* more frequently than the sets generated by the two groups who sorted the writing samples. That is, to a considerable extent, the descriptors base differences in proficiency on students being able to do more or less of something such as "can write on some concrete and familiar topics (Intermediate Low)" and "can write effectively about a variety of topics, both concrete and abstract (Advanced)." Although there is some quantification in the sets of features produced by the groups, the focus is generally on the description of what the group members actually saw in the writing samples such as "emergent paragraph structure, beginnings of relevant ideas present that could be developed (Low Mid, Group 2)."

The categories of *organization, structure,* and *vocabulary* appear across all levels of the descriptors and the groups' features, except the novice range in the descriptors. There are other categories of features, however, that lack the same high degree of consistency. Two categories that are present in the descriptors and not represented in either group's set of features are *writing skill* and *topics/task*. The categories *literacy* and *mechanics* are present only in the descriptors and Group 1's list of features. The category *L1* appears only in the Group 1 list. Overall, for Groups 1 and 2 and the descriptors, there is inconsistency in the use of the categories within levels and ranges. For example, in the descriptors,

vocabulary is not present in Novice Low, Novice Mid, or Superior, but appears in all other subcategories.

In the low range of the descriptors, there is an emphasis on *literacy* and *writing skill* that is not present in the two groups' sets of features. *Structure* and *communicative success* do not appear until Novice High and, more importantly, *organization* does not appear at all at the low level. However, this category is very important to Group 1 and 2 and is specified quite clearly in their sets of features. Another difference between the descriptors and the two sets is that *length* is similarly paced in the two groups' features such that by Low High, students are producing paragraph-level writing. However, in the descriptors, students can only "produce sentences and short phrases" at Novice High.

At the intermediate or mid range, two major differences emerge between the descriptors and the groups' features: (a) the absence of the category *communicative success*; and (b) the lack of development and specificity in the category *organization*. *Communicative success* was an important focus in the two groups' efforts to distinguish between papers; however, it is totally absent in the descriptors. Likewise, *organization* is a critical distinguishing feature for the two groups and is well-developed in both sets of features. In the descriptors for the intermediate range, however, *organization* is described identically across subcategories making it difficult to use this feature to distinguish papers from one another.

Intermediate Low:	Demonstrates some evidence of organizational ability.
Intermediate Mid:	Is able to organize and provide some support.
Intermediate High:	Shows some ability to write organized and developed text.

Furthermore, in the lists of features for the two groups, focus on and development of theme/topic are important components of *organization*. However, in the descriptors, the category *topics/tasks* instead describes the types of topics and tasks a student can do, but does not address attention to task.

Finally another similarity between the descriptors and the two groups' lists of features at the intermediate range is an apparent difficulty in actually articulating features. In the descriptors, Intermediate Mid is the least developed mid-range subcategory, as is the Mid High category for Group 1 and the Mid Mid

category for Group 2. While there is some measure of agreement between the descriptors and the two groups' sets of features, in that there are three subcategories at the Mid range, there are clearly problems in describing the distinguishing characteristics.

In the high range (including Superior and Distinguished[3]), one of the most notable differences between the descriptors and the sets of features for Groups 1 and 2 is the number of levels. The two groups have two to three levels in the high range while the descriptors have four. In general, the descriptions of features in all three sets tend to be clearer and more specific at this range. However, one exception is that there are few distinguishing differences between the Advanced and Advanced High descriptors except the degree to which the features are evident in a paper or papers. In other words, one must differentiate between papers on the basis of quantity. Another major difference at the high range is the emphasis on audience and purpose in the descriptors. In the sets of features for the two groups, there is an emphasis on addressing the topic, but no mention of audience.

The findings from this comparison of the descriptors to the features generated by the two groups have raised specific issues regarding the language of the descriptors and the number of levels. These results, in addition to the lists of articulated features created by the ESL expert writing groups, should prove useful for refining the descriptors.

Working Group Subcommittee Sort with Original Descriptors

A subcommittee of the working group, one representative from each segment, attended a one-day meeting with CSE staff to assign writing samples to the descriptor levels (see Appendix F for the original descriptors). Since the working group members developed the descriptors, CSE staff felt valuable feedback would be obtained by having them assign writing samples to the levels they had defined. Fourteen samples representing a wide range of ability and each of the segments were selected from the group of papers read during the expert sorts described in the preceding section. The subcommittee read the set of

[3] The distinguished level was included by the descriptor developers in recognition of there being non-native speakers as well as native speakers who reach a higher level of proficiency than would be represented in either an advanced ESL or mainstream English class.

samples, independently indicated the appropriate level, and made notes regarding problems in using the descriptors to separate the samples.

Although the task was not a "rating" task per se, going through the process of placing a paper into one of the descriptor levels is very similar to assigning a rating or score to a writing sample. As it turned out, the task was difficult for the group. There was no agreement on the placement of any one paper, and there were only four papers for which there was a majority agreement (3, 1). The group felt that all the mid range papers were difficult to place in part due to a lack of completeness about grammar and development in the descriptors. In addition, several other concerns about the descriptors surfaced, namely—issues dealing with descriptor use, the language of the descriptors, the organization of descriptor statements, and degree of specificity. A discussion of each one follows.

Descriptor use. The group experienced difficulty in trying to use the descriptors for the purpose of placing a single sample at a level. They felt that the descriptors are too general to be used for "ranking" a paper and should not be used as a scoring protocol; the descriptors characterize a writer, not a single paper, and thus seem to be most appropriate for use with a portfolio. Specifically, for most levels, the first bullet describes the range of writing types a student can produce. For example, at Intermediate Mid, the first bullet states *can write on some concrete and familiar topics*. For both Advanced High and Superior, bullet three refers to the writer as being able to *tailor writing to purpose and audience*. Multiple samples are needed to adequately assess the writer's ability in this regard.

Descriptor language. The language of the descriptors also presented problems. Specific issues emerged regarding wording. For example, the relative amount captured by "limited" in *demonstrates limited control of sentence structure and punctuation to indicate sentence boundaries* (bullet three, Intermediate Mid) is not clearly differentiated from "some" in *displays some control of sentence structure and punctuation to indicate sentence boundaries, but often makes errors* (bullet four, Intermediate High). In Novice High, bullet four, one must determine if a writer is producing sentences and short phrases which have been *previously learned*. However, how does one differentiate between "previously learned" material and that which is not?

In Advanced High, the second bullet (*displays rhetorically effective organization and development*) and sixth bullet (*uses a variety of sentence structures for stylistic purposes*) are possibly redundant. Further, the group felt it was difficult to infer that the presence of a variety of sentence structures in a sample meant that the writer used them intentionally for stylistic purposes. As mentioned above in the comparison of the descriptors and the expert levels, the subcommittee members also found it difficult to distinguish between bullet two, Intermediate Mid (*is able to organize and provide some support*) and bullet two, Intermediate High (*shows some ability to write organized and developed text*). In addition, they felt that bullet two in Intermediate High is too low for that level.

Descriptor organization and emphasis on critical level features. As indicated in the example above, the group articulated a need to revisit the organization and ordering of descriptors within and across levels. Some bullets seem better placed in different levels. For example, the group felt that bullet four in Intermediate Low, *demonstrates some evidence of organizational ability*, should be placed higher within the descriptor because it is an important feature at that level. This implies that the order of the bullets in the descriptors may need to be adjusted to better reflect what is most critical at each level. Along these lines, the group raised the issue of how to judge samples in which writers "attempt" to use more challenging vocabulary and structures with varying degrees of success. In other words, how should the student's willingness to take risks factor into judgments about the writing? The descriptors tend to note limitations of writers, not what they can do or attempt to do.

Limitations in descriptor specificity. The group identified several issues, including the role of sample length, response to topic, maturity of thought, and register, that they felt are not adequately addressed. Some descriptors are underspecified; for example, grammar at Intermediate Mid, in particular, is too vague. In fact, the group felt that the intermediate range is difficult to use in general because there is not enough detail built into the descriptors about development and grammar.

Another issue raised by the subcommittee regarding the lack of specificity in the descriptors is related to attention to task. It seemed to them that this feature became more important between the intermediate and advanced ranges, but the shift is not specified in the descriptors. There were writing samples that led them to consider the question: If the goal of responding to topic is put aside, could a

paper be assigned to a higher level? Another related consideration is the question of how papers should be evaluated, especially at the more advanced levels, if audience and purpose are not specified in the task. Thus, the relationship between task and the descriptors is an important one.

The working group subcommittee helped to verify important problem areas within the current writing descriptors that will require clarification and revision in the future. These problem areas have already been identified in part through the comparison of the descriptors to the products of the ESL writing expert sort. Issues related to purpose or use of the descriptors and the language of the descriptors will be revisited later in this report.

Expert and Non-expert End-user Sort

To examine the use of the descriptors for specific purposes by experts and non-experts, a user sort was conducted in which a subset of the samples sorted by the ESL experts was mailed to eleven potential users across segments. The packets included instructions, descriptors, writing samples, a descriptor placement worksheet, and a questionnaire. The participants were instructed to take two to four hours to complete the tasks and then to return the entire packet by mail. Ten packets were completed and returned to CSE. Brief phone interviews were conducted with seven of the participants afterwards.

Using the current writing descriptors, the participants assigned the samples to descriptor levels and completed a short questionnaire (see Appendix G). Users were asked questions such as the following:

1. How might you envision these descriptors being used at your institution? Who do you think would or should use them?

2. Were the descriptors easy or difficult for you to use?

Characteristics of the end users. Five of the participants were unfamiliar with the descriptors, two had heard of them, and three were very familiar (e.g., one of the three has used the descriptors in graduate courses on methods and curriculum design). Seven of the participants work with ESL students frequently as a part of their job responsibilities, including helping students with career and educational planning, and classroom instruction. Two participants work with students infrequently although they are responsible for program design and other similar duties that directly affect ESL students. The tenth participant is a

counseling assistant who occasionally works with ESL students who are communications studies and education majors. Participants' decision-making responsibilities regarding ESL students include school program management, instruction, evaluation, curriculum, and hiring qualified instructors.

Descriptor uses. The most frequently named potential use for the writing descriptors was placement (five of the ten participants). The following potential uses were each named once: classroom assessment, program evaluation, promotion, admissions, and as a reference document. Two participants thought use of the descriptors would be problematic and, for that reason, did not state any potential uses.

All but two of the participants thought descriptors for the other skill areas would be useful; in particular, for reading and speaking. No one specifically mentioned listening. The participants felt the descriptors would be useful for testing, admissions, curriculum, articulation, and to clarify program goals. Two of the participants stated that the descriptors should only be used if users were trained in how to apply them. One person felt they were of limited use and need more definition, and another stated that she has little need for them since her campus does not have an ESL program.

Six of the participants said they had problems assigning the writing samples to the descriptors, three felt they were easy to use, and one person felt that some descriptors were easier to use than others. Reasons given for their difficulties include: (a) the descriptors are not explicit enough; (b) definitions are needed for some of the terms used in the descriptors; (c) differences between the levels need to be more clearly delineated; and (d) a single writing sample is not sufficient for placement into levels above Intermediate High.

Participants comments regarding the descriptor levels and ranges include: (a) it was difficult to place papers into the Novice Low and Novice Mid levels; (b) indicators are generally weak at the intermediate range; (c) Intermediate Mid and Intermediate High are the most difficult levels to use; and (d) more than one sample is needed from each student to use levels above Intermediate High.

Phone-interview responses. Seven participants were asked two questions during a brief phone interview which was conducted after the packets were returned:

1. Do you have any recommendations for refining the descriptors to make them more user friendly?

2. Could you comment on the adequacy of the number of levels for use in your segment?

When asked the first question, most participants stated that they did not have much to add to what they had already noted on the end-user participant questionnaire. However, three of the seven participants interviewed strongly believe that the descriptors should be more specific, particularly if they are to be used with a single task. Two others asked which features take priority at different levels and if the priorities for these features vary from level to level.

When asked the second question, six of the participants felt that the current number of descriptor levels is adequate and necessary to address the range of students in their segment and others. One of the six felt initially that there were too many descriptors, but after assigning papers to them he found the number of levels to be adequate. The seventh person felt overwhelmed by the number of descriptors and did not feel that the distinctions between the levels were clear.

Assignment of descriptor levels to writing samples. The participants assigned descriptor levels to fourteen writing samples that had already been sorted into levels by the ESL writing experts and assigned to descriptor levels by the subcommittee members. No training was provided regarding the use of the descriptors. They were each given a Descriptor Worksheet (see Appendix H) on which they could note their placement of each paper and any comments they had on using the descriptors for this purpose.

There was close agreement on the assignment of several of the papers to the descriptor levels. In fact, when the end-user placements are combined with the subcommittee member placements, six papers emerge with very strong tendencies toward placement into a single descriptor level (5 to 7 participants of 14 in exact agreement and 4 to 8 participants with adjacent placements for these six papers). More exact placements might have been obtained if the users had received training. These six papers fell into the Novice High, Intermediate Low, Intermediate High, and Advanced categories (see Appendix I for the fourteen writing samples).

Participants expressed difficulty using some of the ranges, the intermediate range in particular. Placement of four of the six papers in the intermediate range

are problematic because they cluster on either Intermediate Low or High, but never on Intermediate Mid (e.g., 5 IL, 2 IM, 6 IH, and 1 A for paper #106). There seemed to be difficulty placing papers that fell into the advanced range and above. Placement of the papers into these levels spread out, almost equally in some cases, among the high level descriptors (e.g., 3 IH, 4 A, 3 AH, and 4 S for paper #586). Papers that fell into the novice, novice-intermediate, and intermediate-advanced ranges clustered logically around a single level with declining numbers of adjacent papers on each side (e.g., 3 IM, 5 IH, 5 A, 1 S for paper #306). These are the same levels with the highest degree of agreement among the participants regarding placement of papers. The high agreement may indicate that Novice High, Intermediate Low, Intermediate High, and Advanced are currently the most clearly defined levels.

Participants' comments on the Descriptor Worksheet generally indicated the features of the papers and/or descriptors used to place the papers into levels. There were also several comments regarding the short length of three papers and how difficult the reduced length made it to use the descriptors to place these papers. One participant stated that length deserved mention although including the feature in the descriptors could "muddy the reader's evaluation." One participant questioned whether "integrating source material," which appears in the advanced range and above, is the same as "uses appropriate examples" (in samples that are obtained in this way, e.g., timed responses) in the low and intermediate ranges.

Other comments included: (a) it is difficult to place papers in the intermediate range; (b) it is difficult to distinguish between Intermediate High and Advanced; (c) there seems to be a big step between Intermediate High and Advanced; and (d) terminology needs to be defined (e.g., concrete, familiar, personal).

Suggestions for Next Steps

The findings of this study clearly indicate a range of needs by professionals who work with ESL students for a tool such as the descriptors. Instructors view the descriptors as an instrument for classroom instruction and curriculum development. Career and educational counselors think of them in terms of admissions criteria, articulation, and progress indicators for goal setting, particularly with regard to the students' need to perform at native-speaker levels

in order to take mainstream courses, graduate, and so forth. Program managers think of the descriptors in terms of assessment needs, curriculum, and articulation between schools and segments. All of these expressed uses are important and should be acknowledged; however, one set of descriptors cannot validly fulfill so many needs at once. Furthermore, because the descriptors were developed to capture proficiency levels in a general sense, the levels are not currently defined in enough detail for application across the range of uses identified in this study. Thus important decisions and revisions should be made prior to the final validation of the writing descriptors and the validation of the descriptors for the other skill areas. On the basis of the results of this study, suggested decision-making and revision guidelines for next steps are provided below.

Considerations for Revision of the Writing Descriptors

Three major considerations critical to the continued validation of the writing descriptors emerged from this study: range of descriptor uses, refinement of descriptor levels, and clarity in descriptor content, including organization and language. These considerations will be critical in the validation of the other skill areas as well. Each is discussed in turn followed by suggested next steps for validation of the writing descriptors and suggested steps for validation of the descriptors for the other skill areas.

Descriptor uses. Critical to the validation of the writing descriptors and the validation of the descriptors for the other skill areas is determining for whom the descriptors are intended and how they will be used. An issue that was raised repeatedly in this study was the question of how the current descriptors can be used to "rank" a single paper. As they are written, the uses are limited to and only appropriate for classroom assessment, curriculum development, and possibly promotion or exit assessment—situations that provide multiple samples of a students' work such as a student portfolio.

One recommendation regarding descriptor uses involves validating the descriptors for two or three specific, critical uses. The most critical uses identified in this study are: articulation within and across segments, classroom assessment, curriculum development, and placement testing. Although changes to the descriptors are needed to assure their effective use, they are already oriented towards describing a student rather than a writing sample; thus it seems

appropriate given that focus to begin by validating them for articulation and the two classroom-related uses mentioned above, classroom assessment and curriculum development. Using the descriptors with placement tests will likely involve extensive revisions to the language of the descriptors including reorientation towards a single writing sample. Validating the descriptors for these different uses may result in sets of related but different descriptors and/or different sets of guidelines for use. After determining which uses the descriptors will be validated for, the descriptors should be modified accordingly.

Other descriptor uses will require additional validation considerations. For example, to validate the descriptors for use in counseling situations, the issue of relating a student's English proficiency level to the English proficiency level necessary to take mainstream classes with native speakers of English must be addressed. Counselors should be able to inform and advise students about the level at which they need to perform in order to participate in mainstream classes. Their needs are different from ESL experts in that they view an ESL student's performance with respect to the levels of proficiency necessary to participate in mainstream classes. Thus, designating a target proficiency level vis à vis the descriptors would be necessary. Specific cutoff levels may also be necessary for placement testing, since placement tests may need to differentiate between students who require services and those who do not.

Descriptor levels. Although end users in this study did not use all the descriptors, they were generally comfortable with the current number of levels when they used the descriptors to place papers. However, in earlier phone interviews, some end users indicated that fewer levels are necessary, particularly at the higher ranges (Advanced through Distinguished). This potential need for fewer levels was partially confirmed during the expert sorting process. As noted earlier, one expert group arrived at seven to eight levels, and the other group arrived at eight to nine levels. In both cases, the groups had fewer levels in the advanced range than the descriptors do. This may have been a function of the available writing samples. Regardless, the number of levels present in a set of validated descriptors should represent the number of levels that have been identified empirically as well as theoretically.

Specific recommendations regarding the descriptor levels for writing are to: (a) articulate features of the overarching ranges (novice, intermediate, and advanced) within which more specific descriptors for each range will fall; and (b)

refine the middle and advanced levels. Creating a list of features for these overarching ranges will help with both articulation and curriculum development by simplifying the comparisons between ranges and levels among schools. It will also be easier for users to determine initially which broad range a student or his/her performance falls into; the next step would be to assign a level within that range to the student or the performance.

Attention should also be focused on delineating the differences between levels within a given range, especially with the intermediate and advanced ranges. The intermediate descriptors were problematic for the subcommittee members, the end users, and even for the experts who attempted to articulate their own levels and distinguishing features. Addressing problems described below with the organization and language of the descriptors should facilitate these improvements.

Descriptor content: organization and language. While end users and the subcommittee members were able to reach a surprising level of agreement on assigning six of the fourteen papers to descriptor levels, there were numerous problems assigning the other eight. These problems, along with specific comments made by participants in this study, have led to the following conclusions:

1. More specificity is needed in the descriptors.

Users felt the descriptors often lack specificity in important areas such as grammar, and organization and development. Not only did end users and subcommittee members point to problems with organization and grammar, the comparison of the expert groups' lists of features to the descriptors indicated that organization is a major weakness. Other descriptor specificity issues are the need to define more clearly what is meant by terms, such as *personal, concrete,* and *writing skill* at the low and high ranges, and the need to clarify why certain features appear at some levels and disappear at others (e.g., *writing skill* appears only in the low and high levels, not in the mid levels). Attention to topic or task is a major issue that should be addressed in descriptor revisions. Features that describe a student's level of communicative success are also missing at the intermediate range.

2. Descriptors should describe more and quantify less.

The subcommittee and end users had difficulty placing writing samples into descriptor levels when features in the descriptors were quantified. Particularly in the intermediate range, features of the category *organization* tend to be quantified which makes it more difficult for users to distinguish between levels. The frequent use of quantifiers in the descriptors is likely due in part to the notion of a proficiency continuum which underlies the descriptors. A continuum approach to describing language proficiency may lend itself to a focus on the increase or decrease of observable features, but when the goal is to identify and describe specific points on the continuum, as with the descriptors in this study, referencing more or less of a feature is not sufficient for effective descriptor application. Indeed, end users felt they needed definitions for terms that quantify such as *few* and *some*. The meaning of these terms can be subjective and could thus lead to low reliability in the application of descriptors that include them. However, when features are described, descriptor users have something concrete to look for in a sample.

3. An approach to weighting the language features within the descriptors needs to be specified.

Finally, a determination must be made as to whether the language features within each descriptor should be viewed holistically, and thereby be considered as having equal weight in the application of the descriptor, or whether they should be prioritized within the descriptors. Some features seem to be more important than others in defining specific levels which suggests that an approach which prioritizes the features in some way for purposes of application may be the most appropriate.

Next Steps in the Validation of the Writing Descriptors

Although considerable progress has been made towards validation of the writing descriptors, feedback from the current study, including the need for two orientations for their use—single sample and multiple sample—dictates that the descriptors be modified before validation can be completed. Indeed it seems clear that two sets of writing descriptors are needed to meet the range of needs articulated thus far. The content of the sets would be similar but would vary on specific points such as addressing a topic in a single sample versus demonstrating awareness of audience and purpose across multiple samples from the same writer. The features specified in the two sets of descriptors would be

complementary, differing only to the extent necessary to accommodate the two orientations for descriptor uses.

The procedures for continued validation would be the same for both sets with the exception of an additional step for the multiple-sample descriptors. A brief discussion of the steps to complete the validation process for the two sets of writing descriptors follows, concluding with a summary of the steps at the end of the section.

Validation of single-sample descriptors. To validate the descriptors for use with single samples, such as in placement test development, the descriptors must first be modified. The experts' sets of features and suggestions from this study for clarification of the language and reorganization of the descriptor features can be used in this step. Since the descriptors would be oriented towards single-sample use, features that would require more than one sample such as "can write on a range of topics" must be removed or adapted.

Once the descriptors have been modified, guidelines for the use of the descriptors should be drafted. A small group of ESL writing experts (two from each of the four segments) would then place writing samples into the descriptor levels using the modified descriptors and guidelines. If necessary, further modifications should be made to the descriptors and guidelines based on the results from the ESL writing experts. Next a small-scale user tryout (three users per segment) would be conducted with the writing samples and guidelines used by the experts. This step will inform any final modifications to the descriptors and the guidelines. Finally, a large-scale user tryout (ten users per segment) would conclude the validation.

Validation of multiple-sample descriptors. To validate the descriptors for uses that require multiple samples of student work such as curriculum development, classroom assessment, and promotion or exit assessment, a variety of sample types from the same writer or portfolios must be collected. Experts would then review the samples, sort the collections of samples into levels, and characterize the language features that describe the writers' range of abilities at each level. These empirically-derived features would be compared to the original descriptors to refine their usefulness with multiple samples. Guidelines for using the descriptors should then be drafted.

Next an expert tryout using the revised descriptors and guidelines and a new batch of writing samples or portfolios would be carried out. Modifications should be made as needed. As with the single-sample descriptor validation, a small-scale user tryout, possibly followed by final modifications of the descriptors and guidelines, and a large-scale user tryout would conclude the validation.

The suggested steps for the continued validation of the writing descriptors are summarized in the list below.

- Determine uses: Articulate uses for which the descriptors will be validated.

- Modify descriptors: Modify descriptors for single-sample uses based on the empirically-derived features from the current study. Descriptors can be modified for multiple-sample uses after the two steps described below have been carried out.

- Sample collection: Collect samples as appropriate for intended uses of descriptors. Some unused single samples are already available from 1997-1998 work; however, additional samples may be needed at the higher ranges of ability since few of the samples, even from the mainstream English classes, fall into the Advanced High, Superior, or Distinguished categories.[4] Multiple samples of a range of types from the same students or a portfolio, if available, will also be needed if validating the descriptors for uses involving multiple samples.

- Extra steps for multiple-sample uses: ESL writing experts review the collection of samples for each student, sort the collections into levels, and characterize the language features that describe the writer's range of abilities at each level. These empirically-derived features would then be compared to the original descriptors and modifications made as warranted.

- Descriptor guidelines: Draft guidelines for use of the descriptors. For single-sample uses, it will be important to select anchor papers from the current work.

- Expert tryout with descriptors and guidelines: ESL writing experts place writing samples into levels using the modified descriptors and guidelines (8 experts, 2 from each segment). Refine modified descriptors and guidelines as warranted following use by experts.

[4] Additional writing samples across segments and proficiency levels should be collected on one or two new topics to ensure that the descriptors can be applied effectively regardless of topic.

- Small-scale user tryout: Conduct small-scale user tryout (12 users, 3 per segment). Make final revisions to descriptors and guidelines based on feedback from users.

- Large-scale user validation tryout: Conduct large-scale user tryout with final version of the descriptors and user guidelines (40 users, 10 per segment).

Validation of the Speaking Descriptors

To validate the speaking descriptors, the uses for which the descriptors will be validated must be clarified (e.g., test development, curriculum, classroom applications) and the descriptors modified accordingly based on empirical evidence. For example, if they are to be validated for uses that involve a single sample of student performance, the language of the descriptors should focus on describing features found in a single sample (e.g., addressing the topic) as opposed to multiple samples (e.g., adjusting to different audiences). As they are currently written, multiple samples of student work are needed to use the descriptors for all the skill areas.

Whether the descriptors are validated for use to evaluate a single sample or to characterize the ability of an ESL student, the validation process would ideally include classroom observations leading to the identification of typical oral tasks. Tasks would be selected for their potential in eliciting a range of performance from students in all four segments. The most promising tasks would then be piloted for use in the validation. The best tasks should be selected from the tryout and adapted if necessary for the larger data collection efforts. Data should be collected on tape from a range of proficiency levels across the four segments. However, since the data collection for this skill area will be more time consuming than for writing, it is likely that fewer samples will be collected. If the descriptors are to be validated for uses that involve multiple samples, samples from the same speaker, across tasks, must be collected and then later sorted by experts. This could be done either in one sitting or across time.

After the data are collected, steps similar to those in the writing validation can be followed. First an expert sort should be conducted during which ESL oral language experts articulate the features of the samples that they have sorted into levels. This step should be followed by a comparison of the features identified by the experts to the speaking descriptors; modifications of the descriptors based upon the results should be carried out. Guidelines for use should be drafted, and

then a new set of samples assigned to levels using the modified descriptors and the guidelines. Modifications as warranted would be made and a small-scale user tryout conducted. Final revisions to the descriptors and guidelines would follow. The final step would be a large-scale tryout with intended descriptor users.

The suggested steps for the validation of the speaking descriptors are summarized in the list below.

- Determine uses: Determine uses for which the descriptors will be validated.
- Identify tasks to be used in the validation: Conduct classroom observations to identify typical oral tasks and select potential tasks to pilot. Conduct tryouts of promising tasks and select tasks that elicit the best speech samples for the larger data collection effort.
- Sample collection: Collect speech samples across the range of proficiency levels and across all four segments.
- Expert sort: Conduct expert sort to empirically derive features. Special equipment issues must be addressed when planning this stage, such as the possible need for a language lab, cassette players with earphones, etc. Compare results of sort to the original descriptors and modify as necessary.
- Descriptor guidelines: Draft guidelines for use of the descriptors.
- Expert tryout using descriptors and guidelines: A new set of samples should be assigned to levels by a different group of speaking experts using the modified descriptors and guidelines. The descriptors and guidelines should be modified as warranted following use by experts.
- Small-scale user tryout: Conduct small-scale user tryout. Make final revisions to descriptors and guidelines based on feedback from users.
- Large-scale user validation tryout: Conduct large-scale user tryout with final version of the descriptors and user guidelines.

Validation of the Listening and Reading Descriptors

The procedures for validating the listening and reading descriptors would differ from the validation process described for speaking and writing because ability in these skill areas can best be captured indirectly. The validation procedure will involve establishing an empirical basis for the descriptors by

examining what kinds of listening or reading tasks students can and cannot perform, anchoring the descriptors with level-specific composites of features of performance, and finally validating the descriptors.

First, as with the writing and speaking descriptors, the purposes for which the descriptors will be used must be clarified. The orientation of the descriptors towards either describing student performance through a single event or across multiple events will dictate the types of composites constructed and modifications to the descriptors. Either way, a variety of listening and reading tasks from all four segments across a range of difficulty, including both ESL and mainstream classes, should be identified, ideally through classroom observations. At the same time, a teacher survey would be conducted in which teachers from those classes are asked to identify and categorize typical listening and reading tasks according to difficulty. They would also be asked to identify level-specific features of performance within tasks.

Using the list of identified task types, students would be interviewed about the difficulty of the tasks and asked to indicate which tasks they can and cannot perform. Next student test scores would be obtained for the students interviewed to cross check their interview responses regarding task difficulty. In an ideal situation, available tests for listening and reading would be analyzed, selected, and administered by researchers conducting the study.

Then, through review of available sources—analyses of the tasks, teacher surveys, student interviews, student test scores, and test analysis data—listening and reading experts from each segment would identify features for the range of proficiency levels present in the data. These results would be compared to the existing listening and reading descriptors. The descriptors would be modified on the basis of these comparisons and the composites of the level-specific features of performance would anchor each descriptor level.

Suggested steps for anchoring the listening and reading descriptors are listed below.

- Determine uses: Determine how the descriptors will be used.
- Identify tasks: Conduct classroom observations to identify typical listening and reading tasks.

- Identify task difficulty—teacher perspective: Obtain teacher feedback regarding typical performance levels on tasks and characterization of the tasks in terms of difficulty.

- Identify task difficulty—student perspective: Conduct student interviews regarding difficulty of tasks.

- Select listening and reading tests: Choose tests that include tasks identified as relevant through the classroom observations.

- Collect and analyze data: Administer test to students and analyze scores. Analyses of the tasks should also be compiled.

- Anchor the descriptors to composites: Listening and reading experts will build composites of level-specific performance features based on task analyses, teacher surveys regarding tasks, analyses of student interviews, test scores, and test analysis data. The composites will be compared to the existing descriptors. Modifications will be made on the basis of these comparisons, and the composites will anchor the descriptors.

- Descriptor guidelines: Prepare guidelines for use of the descriptors.

Alternative Data Collection Methods

Since the validation methods suggested in this report can be time consuming and expensive, alternative data collection methods or validation approaches should be considered. For example, although the skills are defined in isolation from one another in the descriptors, it may still be possible to use an integrated skills approach to collect data. The collection of reading and speaking data could be combined, i.e., students could complete a series of reading tasks followed by short controlled speaking tasks based on the reading with an interviewer or another student. These data would be recorded and procedures similar to the ones outlined in the preceding section could be followed. Another approach to collecting data would involve identification of a range of multi-skills classrooms across the state that would agree to participate in the validation study for an entire quarter or semester. Data for all skills, in addition to task and student information, could be collected in these classrooms and compiled for use during expert sorts.

Descriptor Handbook

After a final validation has been carried out for the writing descriptors and descriptors for the other skill areas, the revised descriptors should be released with the guidelines for their use in a Descriptor Handbook. The descriptor handbook would function as both a user guide and training handbook for inexperienced users and would ensure more valid application of the descriptors. The Handbook would explain appropriate situations for descriptor use and procedures for their application. It should also include samples of student work that anchor each descriptor level.

Final Comments

This validation study has raised many important issues regarding use of the writing descriptors. It has also helped to pinpoint problem areas in the language and organization of the descriptors. Based on the results of the study, suggestions for next steps have been made which include key decisions about the purposes for which the descriptors will be validated and specific areas in the descriptors that require modification.

This study has also resulted in the development of a process which can be applied to the final validation of the writing descriptors or adapted for the validation of descriptors from the other skill areas. Whichever approaches to validation are used, a critical first step in the continuation of this work will be clear specification of descriptor uses. Once a validation process has been carried out for a specific use, caveats should be issued with the release of the validated descriptors emphasizing that the descriptors have only been validated for that use and may not be valid for others. Guidelines for use, along with anchor papers or language samples, must be linked to the validated descriptors. These steps may not prevent incorrect or inappropriate use of the descriptors, but they will help to inform those for whom the descriptors are intended.

References

Butler, F. B., & Stevens, R. (1997, September). *Validation plan for the second language proficiency descriptors for public high schools, colleges, and universities in California*. Center for the Study of Evaluation, Graduate School of Education & Information Studies, University of California, Los Angeles.

Mlynarczyk, R. (1991, Autumn). Is there a difference between personal and academic writing? *TESOL Journal, 1*(1), 17-20.

Upshur, J. A., & Turner, C. E. (1995, January). Constructing rating scales for second language tests. *ELT Journal, 49*(1), 3-12.

Appendix A

Working Group Members 1997-1998

Gari Browning, Project Director
Coordinator, Institutional Research
Orange Coast College

Donna Brinton
Academic Coordinator, ESL Services Courses
University of California at Los Angeles

Roberta Ching
ESL Coordinator, Learning Skills Center
California State University, Sacramento

Robert Dees
President, English Council of California & Two-year Colleges
Dean, Literature & Languages, Orange Coast College

Susan Dunlap
Program Assistant, Bilingual/ELD Staff Development
West Contra Costa Unified School District

Melinda Erickson
Lecturer, College Writing Programs
University of California at Berkeley

Katheryn Garlow
Past President, California Teachers of English
to Speakers of Other Languages
Associate Professor in English in a Second Language
Palomar College

Margaret Manson
Dean of Instruction
Orange Campus
Rancho Santiago College

Deborah Poole
Coordinator, ESL Writing Courses
San Diego State University

Linda Sasser
Secondary ESL Program Specialist
Alhambra School District

Appendix B

End-user Interview Participants: Job Titles and Schools

CSU Bakersfield
Director of the Writing Program

CSU Northridge
Professor, English Department

James Monroe High School
Bilingual Site Coordinator

Palomar College (CC)
Counselor

Pinole Valley High School
ESL Department Chairperson

Santiago Canyon College (CC)
Director of Student Services

UC Berkeley
Career Guidance Counselor

UC Berkeley
Assistant Director
Graduate School Instructor Teaching and Resource Center

UCLA
Associate Director
Undergraduate Admissions and Relations with Schools

Appendix C

End-user Interview Protocol

General Introductory Questions

1) Have you heard of the document called California Pathways? *If no, explain briefly. If yes, ask if the person is using it and for what purpose.*

2) As part of your responsibilities, how often do you work with students who speak English as a second language?

3) What kinds of decision-making responsibilities do you have regarding ESL students?

4) How does your institution identify ESL students? *What assessment do they use, if any?*

5) What support services does your institution provide to ESL students?

Placement

6) If you place ESL students, how do you place them? *If no, do you know who is responsible for placing them?*

7) Are there any problems at your school regarding student placement?

Guidance

8) If you provide guidance to ESL students, what kinds of information about ESL students would be useful to you? *If no, do you know someone who does provide guidance or counseling to ESL students?*

9) When providing guidance to ESL students who plan to transfer to another college (or go on to college from high school), do you encounter any problems (such as not being able to link coursework at one school to course requirements at another school)?

Assessment/judgments of student proficiency

10) Do you make judgments about students' English proficiency? *If so, which skill (s) do you make those judgments about? If not, who at your institution makes these types of judgments?*

11) On what basis do you make those judgments?

Faculty issues

12) Do you interact with faculty who work with ESL students?

13) If faculty need to make recommendations or judgments regarding ESL students, do you know how they do that (on what basis)?

Descriptor questions

14) If you had a tool or instrument that you could use to judge or discuss ESL student proficiency levels, how many levels of proficiency do you think you would use (or need)? *How specific would you need it to be?*

15) Would it be useful to you to have a set of language descriptors for student proficiency levels in reading, writing, speaking, and listening? *How would you use them? If no, can you think of anyone who could use them?*

16) Would you use all the skill areas (reading, writing, speaking, listening)? *Which ones would be most important to you?*

17) Can you think of any other issues regarding ESL students that you feel could be addressed by having a set of language proficiency descriptors (e.g., curriculum development, articulation between different campuses)? *Write in any misc. comments from interview.*

18) What are some problems ESL students have at your institution?

19) What are some problems at your institution regarding ESL students?

Appendix D

Language Features for Levels Identified through ESL Writing Expert Sort for Groups 1 and 2

Reorganized Versions

The category headings (in italics) were generated by CSE staff to systematize features across the two groups.

Language Features for Levels Identified in Writing Samples
Expert Group 1
Topic A: Teenage Difficulties

LOW

- *Communicative success:* Some intelligible sentences
- *L1:* Uses primary language, syntax may reflect primary language
- *Length:* Words to sentences, fragments, phrases
- *Literacy:* May not show grasp of sound/symbol correspondence
- *Mechanics:* Errors in mechanics
- *Organization:* Some attempt at development, topic sentence and cohesive paragraph level writing, awareness of topic
- *Structure:* Errors in grammar, simple sentences, no complex sentences
- *Vocabulary:* Key vocabulary, inventive spelling

(1) Low Low -
Communicative success: May be unintelligible
Length: Very brief; may be few if any clauses, sentences; an attempt at words
Organization: Pre-paragraph, lacks evidence of paragraph

(2) Low Mid -
Communicative success: Errors impede understanding
L1: May have LI influence
Length: Scattered sentences, pre-paragraph
Mechanics: Mechanical errors (e.g., spelling, margins, punctuation)
Organization: Little or no evidence of organization, awareness of topic, may be on topic
Structure: Simple sentences, emerging syntax, boundaries may be unclear; many distracting errors in grammar, grammar limited
Vocabulary: Many distracting lexical errors, very limited

Expert Group 1 language features (continued)

(3) Low High -

Communicative success: Errors interfere with understanding
Length: Paragraph, pre-composition
Mechanics: Many mechanical errors that impede understanding
Organization: Clear, but possibly limited grouping of ideas, no topic sentences appear
Structure: Frequent syntactic errors; many grammatical errors that impede understanding
Vocabulary: Limited vocabulary, errors in use

MID

- *Length:* Multiple paragraphs
- *Organization:* Shows evidence of organization and development of theme/topic, attempt at or some general supporting examples (e.g., facts, details, incidents), topic sentences, able to apply conventions of an essay
- *Structure:* General control of basic sentences (has subject/verb), attempts complex sentences with limited success
- *Vocabulary:* Shows expanding vocabulary and alternate word choice, demonstrates and experiments

(4) Mid Low -

Communicative success: Expresses self despite vocabulary limitations
Length: Multiple paragraphs
Organization: Most development relevant to topic, may have topic sentence, paragraph unity clear with awareness of topic; development-some details and facts that may not be tied to topic; explicit control of organization
Structure: Controls basic sentence patterns, may attempt complex sentences (e.g., adjective clauses, parallel structures), awareness of form
Vocabulary: Limitations in vocabulary

(5) Mid Mid -

Communicative success: Local and global grammatical errors exist but do not prevent comprehension, wordiness or redundancy [evidence of circumlocution]
Organization: Clear organization, ideas clearly expressed, support present and varied but may be limited or general
Structure: Sentence variety, controls simple sentence structure, some control of complex and multi-clausal sentences; not many grammatical form errors (e.g., -ing instead of -ed in verbs, gerunds for infinitives), local and global grammatical errors but may be fewer
Vocabulary: Limited vocabulary may result in repetitiveness

(6?) Mid High -

(The group was not sure about this level; no papers were actually placed here. There were only "fence sitters.")
Communicative success: Numerous errors occur but tend to be localized; innovative
Organization: Organization good, apparent analysis, emerging focus, development good but may be superficial or general, many relevant examples
Structure: Variety of sentence patterns though there may be some repetitiveness

Expert Group 1 language features (continued)

HIGH

- *Communicative success:* Sophistication of errors high, minimal distracting language errors; fluent
- *Mechanics:* Control of mechanics
- *Organization:* Focused, well organized and developed, ample and relevant specific support/examples
- *Structure:* Good syntax, variety of sentence structure with good subordination and transitions; control of grammar and structure
- *Vocabulary:* Appropriate word choice (synonyms and nuances) and use of idiomatic language

(7) High Low -
Communicative success: Less ambitious, non-distracting errors
Organization: Unified or organized and developed, addresses topic but focus may drift, elaboration of ideas present
Structure: Variety of sentences, errors in syntax (ESL markers) but not distracting; verb tenses mastered (few errors)
Vocabulary: Variety of vocabulary, limited/controlled use of vocabulary, not distracting despite some ESL markers

(8) High Mid -
Communicative success: Variability, engagement, "flair", takes risks, apparent effort to use sophisticated thought, ESL markers (sentence/mechanics)
Mechanics: ESL markers in spelling and mechanics
Organization: Easy to follow, well organized and focused, fluid, not choppy, has transitions, readable
Structure: Apparent effort to use complex syntax (conjunction, subordination), ESL markers
Vocabulary: Expanded vocabulary, high level vocabulary appears, apparent effort to use sophisticated lexicon (e.g., synonyms, nuances, and idiomatic language), ESL markers (esp. spelling)

(8/9) High High -
Communicative success: A few minor local errors like a NS might make, no errors that impact meaning, no ESL markers (approaches NS, NS-like)
Mechanics: Controls mechanics, some NS-like errors in spelling and mechanics
Organization: Well developed, focus tight with specifics and examples
Structure: Controls grammar, sentence structure
Vocabulary: Controls vocabulary, idioms

Level Characteristics in Writing Samples
Expert Group 2
Topic B: Discuss Two People

LOW

- *Length:* Short, .69 pages typical length for B papers, .66 for A papers
- *Organization:* Undeveloped
- *Structure:* Coordination, simple syntax—does not attempt anything beyond, tangled syntax, high degree of error in syntax
- *Vocabulary:* Simple/inaccurate vocabulary

(1) Low Low: *Communicative success:* Incomprehensible, attempted to respond
Organization: May not respond to or develop topic at all

(2) Low Mid: *Communicative success:* Uneven comprehensibility
Organization: Emergent paragraph structure, beginnings of relevant ideas present that could be developed

(3) Low High: *Communicative success:* Minimally comprehensible
Organization: Emergent essay structure, may attempt specific examples

MID

- *Communicative success:* Frequent errors
- *Length:* 1.98 pages typical length for B paper, 1.71 for A papers
- *Organization:* Aware of essay structure, some organization, stays on topic most of the time
- *Structure:* May lack cohesion (choppy), ideas not linked
- *Vocabulary:* Limited (colloquial, unsophisticated) vocabulary

(4) Mid Low: *Communicative success:* Errors interfere with comprehension
Organization: Examples present but not integrated, tendency to lose focus
Structure: Errors in syntax (often interfere with comprehension)
Vocabulary: Errors in vocabulary (often interfere with comprehension)

(5?) Mid Mid: (The group was unsure that this level exists. They added it after reading the second batch of papers from Group 1 and had not finished articulating the features.)

Communicative success: Errors sometimes interfere with comprehension
Organization: May lose focus, examples tend to remain general and are not necessarily integrated

(6) Mid High: *Communicative success:* Errors in syntax and vocabulary rarely interfere with comprehension
Organization: Examples better integrated into essay, usually consistent in focus
Structure: Errors in syntax (rarely interfere)
Vocabulary: Errors in vocabulary (rarely interfere)

Expert Group 2 language features (continued)

HIGH

- *Communicative success:* Engages reader, thoughtful, errors do not obscure meaning
- *Length:* 2.35 pages typical length of B papers, 2.3 for A papers
- *Structure:* Occasional errors in syntax that do not obscure meaning
- *Vocabulary:* Occasional vocabulary errors that do not obscure meaning

(7) High Low:
Communicative success: Ambitious
Organization: Minor inconsistencies in focus, examples occasionally not fully developed
Structure: Ambitious syntax, may be misused; minor structure problems
Vocabulary: Ambitious vocabulary, may be misspelled or misused

(8) High High:
Communicative success: High reader engagement
Length: 2.5+ pages
Organization: Compelling examples, well-drafted essay, flows, extensive development, clear consistent focus, clear voice (writes with authority)
Structure: Transitions-ideas clearly linked, transparent structure, native-like syntax, varied and complex

Appendix E

Second Language Proficiency Descriptors
Writing

Reorganized Versions

NOVICE-LOW
- Length: Is sometimes able to write isolated words and/or common phrases
- Writing skill: Has little or no practical writing skills in English

NOVICE-MID
- Length: Can write some familiar numbers, letters, and words
- Literacy: Demonstrates limited awareness of sound/letter correspondence
- Mechanics: Demonstrates limited awareness of mechanics
- Topics/tasks: Can fill in a simple form with basic biographical information
- Writing skill: Has minimal practical writing skill in English

NOVICE-HIGH
- Communicative success: Has limited independent expression
- Length: Can produce sentences and short phrases which have been previously learned
- Literacy: Demonstrates some awareness of sound/letter correspondence
- Mechanics: Demonstrates some awareness of mechanics
- Structure: Uses simple sentence structure, often characterized by errors
- Vocabulary: Uses simple vocabulary, often characterized by errors
- Writing skill: Has some practical writing skill in English

INTERMEDIATE-LOW
- Length: Can write original short texts using familiar vocabulary and structures
- Mechanics: Often exhibits a lack of control over punctuation and spelling
- Organization: Demonstrates some evidence of organizational ability
- Structure: Often exhibits a lack of control over grammar; can write original short texts using familiar structures
- Topics/tasks: Can write on some concrete and familiar topics
- Vocabulary: Often exhibits a lack of control over vocabulary; can write original short texts using familiar vocabulary

INTERMEDIATE-MID
- Mechanics: Demonstrates limited control of punctuation to indicate sentence boundaries
- Organization: Is able to organize and provide some support
- Structure: Demonstrates limited control of sentence structure
- Topics/tasks: Can write on a variety of concrete and familiar topics
- Vocabulary: Often uses inappropriate vocabulary or word forms

INTERMEDIATE-HIGH
- Mechanics: Displays some control of punctuation to indicate sentence boundaries, but often makes errors
- Organization: Shows some ability to write organized and developed text
- Structure: Uses some cohesive devices appropriately; displays some control of sentence structure, but often makes errors
- Topics/tasks: Can write about topics relating to personal interests and special fields of competence
- Vocabulary: Sometimes uses inappropriate vocabulary and word forms

Second Language Proficiency Writing Descriptors, reorganized version (continued)

ADVANCED

- Communicative success: Errors rarely interfere with communication
- Mechanics: Makes some errors in punctuation (but they rarely interfere with communication)
- Organization: Displays clear organization and development; displays an awareness of audience and purpose; demonstrates an ability to integrate source material
- Structure: Uses cohesive devices effectively; controls most kinds of sentence structure; makes some errors in grammar (but they rarely interfere with communication)
- Topics/tasks: Can write effectively about a variety of topics, both concrete and abstract
- Vocabulary: Makes some errors in vocabulary (but they rarely interfere with communication)

ADVANCED-HIGH

- Communicative success: Makes some errors that do not interfere with effective communication
- Mechanics: Makes some errors in punctuation (but they do not interfere with effective communication)
- Organization: Displays rhetorically effective organization and development; demonstrates an ability to tailor writing to purpose and audience; demonstrates some ability to integrate source material
- Structure: Uses a range of cohesive devices effectively; uses a variety of sentence structures for stylistic purposes; makes some errors in grammar (but they do not interfere with effective communication)
- Topics/tasks: Can write about a variety of topics, both concrete and abstract, with precision and detail
- Vocabulary: Makes some errors in vocabulary, but they do not interfere with effective communication

SUPERIOR

- Communicative success: Makes only minor or occasional errors, but they do not interfere with communication
- Organization: Displays strong organization and presents hypotheses, arguments, and points of view effectively; consistently tailors writing to purpose and audience; displays control of the conventions of a variety of writing types; can incorporate a variety of source material effectively, using appropriate academic and linguistic conventions
- Structure: Employs a variety of stylistic devices
- Topics/tasks: Writes effectively for formal and informal occasions, including writing on practical, social, academic and professional topics

DISTINGUISHED

- Organization: Can tailor writing to match specific purpose and audience
- Structure: Employs stylistic variation and a wide variety of sentence structure
- Topics/tasks: Writes effectively on virtually any topic
- Vocabulary: Employs sophisticated vocabulary
- Writing skill: Has writing skills essentially indistinguishable from those of a sophisticated, educated native speaker; fully commands the nuances of the language

Appendix F

Second Language Proficiency Descriptors
Writing

Original version

NOVICE-LOW
- has little or no practical writing skills in English
- is sometimes able to write isolated words and/or common phrases

NOVICE-MID
- has minimal practical writing skill in English
- demonstrates limited awareness of sound/letter correspondence and
- mechanics
- can write some familiar numbers, letters, and words
- can fill in a simple form with basic biographical information

NOVICE-HIGH
- has some practical writing skill in English
- has limited independent expression
- demonstrates some awareness of sound/letter correspondence and mechanics
- can produce sentences and short phrases which have been previously learned
- uses simple vocabulary and sentence structure, often characterized by errors

INTERMEDIATE-LOW
- can write on some concrete and familiar topics
- can write original short texts using familiar vocabulary and structures
- often exhibits a lack of control over grammar, vocabulary, punctuation, and spelling
- demonstrates some evidence of organizational ability

INTERMEDIATE-MID
- can write on a variety of concrete and familiar topics
- is able to organize and provide some support
- demonstrates limited control of sentence structure and punctuation to indicate sentence boundaries
- often uses inappropriate vocabulary or word forms

INTERMEDIATE-HIGH
- can write about topics relating to personal interests and special fields of competence
- shows some ability to write organized and developed text
- uses some cohesive devices appropriately
- displays some control of sentence structure and punctuation to indicate sentence boundaries, but often makes errors
- sometimes uses inappropriate vocabulary and word forms

Second Language Proficiency Writing Descriptors, original version (continued)

ADVANCED

- can write effectively about a variety of topics, both concrete and abstract
- displays clear organization and development
- displays an awareness of audience and purpose
- uses cohesive devices effectively
- demonstrates an ability to integrate source material
- controls most kinds of sentence structure
- makes some errors in grammar, vocabulary, and punctuation, but they rarely interfere with communication

ADVANCED-HIGH

- can write about a variety of topics, both concrete and abstract, with precision and detail
- displays rhetorically effective organization and development
- demonstrates an ability to tailor writing to purpose and audience
- uses a range of cohesive devices effectively
- demonstrates some ability to integrate source material
- uses a variety of sentence structures for stylistic purposes
- makes some errors in grammar, vocabulary, and punctuation, but they do not interfere with effective communication

SUPERIOR

- writes effectively for formal and informal occasions, including writing on practical, social, academic, and professional topics
- displays strong organization and presents hypotheses, arguments, and points of view effectively
- consistently tailors writing to purpose and audience
- displays control of the conventions of a variety of writing types
- employs a variety of stylistic devices
- can incorporate a variety of source material effectively, using appropriate academic and linguistic conventions
- makes only minor or occasional errors, but they do not interfere with communication

DISTINGUISHED

- writes effectively on virtually any topic
- employs stylistic variation, sophisticated vocabulary, and a wide variety of sentence structure
- can tailor writing to match specific purpose and audience
- fully commands the nuances of the language
- has writing skills essentially indistinguishable from those of a sophisticated, educated native speaker

Appendix G

End-user Sort Participant Questionnaire

Name: _____

School: _____

Department: _____

Job Title: _____

1) Have you heard of the document *California Pathways*? If yes, what is your experience or familiarity with it? Have you used the descriptors before?

2) As part of your job responsibilities, how often do you work with ESL students?

3) What kinds of decision-making responsibilities do you have regarding ESL students?

4) How might you envision these descriptors being used at your institution? Who do you think would or should use them?

5) Would it be useful to have a set of language descriptors for student proficiency levels in other skills areas, such as reading, speaking, and listening? If yes, how do you think they would be used and by whom?

6) Were the descriptors easy or difficult for you to use?

Appendix H

Descriptor Worksheet

Paper No.	Descriptor Level	Comment on using the writing descriptors with each paper.
082		
106		
226		
246		
306		
330		
426		
486		
502		
562		
586		
674		
762		
840		

Appendix I

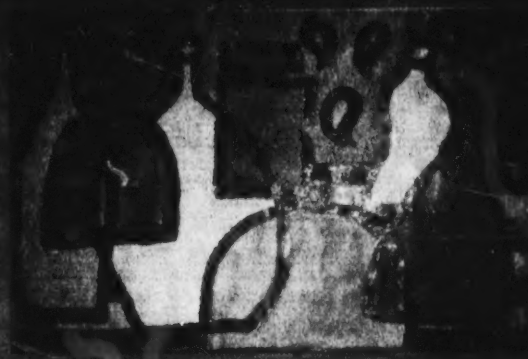
Writing Samples

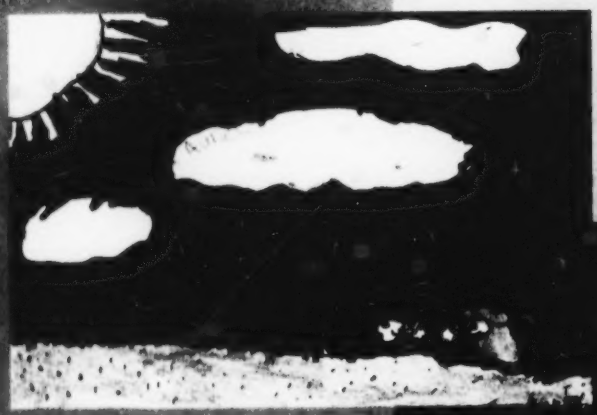
The following fourteen writing samples were used in the expert, the working group subcommittee, and the expert and non-expert end-user sorts.

> Topic B: Write an essay in which you discuss some difficulties that teenagers have growing up. Explain your opinion and give specific examples.

Below is a list of the papers which indicates the segment from which each paper came. Seven of the fourteen papers were identified as exemplars by Group 2. The group's comments about the seven papers are included.

Paper Number	Segment	Comments from Group 2
082 LM exemplar	UC	Beginnings of relevant ideas
106	UC	
226	HS	
246	HS	
306	CC	
330 LH exemplar	CC	Attempts specific examples
426	CSU	
486	CSU	
502 HL exemplar	CSU	No conclusion, but fairly well-developed examples
562 LL exemplar	UC	Quotes prompt
586 HL exemplar	UC	One well developed example; clear focus; mechanical form; sentence-level problems that keep it from HH
674 LH exemplar	HS	Emerging essay structure
762	CC	
840 HH exemplar	CC	Strong; selected because it has ESL markers





BEST COPY AVAILABLE